DOCUMENT RESUME

ED 230 598 TM 830 391

AUTHOR Bradburn, Norman M.
TITLE The Relationship of Federal to Private Statistics.
PUB DATE 29 Oct 82
NOTE 14p.; Paper presented at the Annual Meeting of the Evaluation Research Society (Baltimore, MD, October 28-30, 1982).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Administration; Data Analysis; *Data Collection; *Federal Programs; Financial Policy; *Government Role; Interviews; *Quality Control; *Resource Allocation; Standards; *Statistics
IDENTIFIERS Budget Cuts; *Census Bureau Current Population Survey

ABSTRACT

Federal statistics form the foundation and are the standard for private statistics. Reliable and accurate statistics are increasingly used in complex models of the economy and society that have become important tools for management of large business concerns and for policy planning in governmental bodies. Reductions in government funding threaten the statistical system. These effects undermine assumptions of the quality of federal statistics, which will, in turn, have great negative effects not only on the foundation of the federal statistical system and the Census, but also on private statistics. Further, eliminating the central statistical group in the Office of Management and Budget will result in the loss of coordination in such matters as common definitions and common classificatory schemes that have enormous implications for comparability of data for analysis across surveys and among agencies and for the quality and usefulness of federal statistics. Statistics on society and the economy may be thought of as public goods that further the public interest as well as private interests. Hence they are properly the responsibility of the federal government. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED230598

THE RELATIONSHIP OF FEDERAL TO PRIVATE STATISTICS

by

Norman M. Bradburn

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. M. Bradburn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A talk prepared for the meetings of the
Evaluation Research Society
October 29, 1982
Baltimore Maryland

TM 830 391

THE RELATIONSHIP OF FEDERAL TO PRIVATE STATISTICS

Norman M. Bradburn

Periodically I receive news releases from the ABC News Polling Unit describing results from ABC News/Washington Post polls. Unlike most people who receive this publication, I am more interested in the parts of the release that describe, all too briefly in my opinion, the way in which the poll was conducted and any methodological wrinkles or caveats that would enable the thoughtful reader to understand more completely the meaning of the data presented in the news release. In the latest release, there was a section on general methodology which contained the statement: "Responses were weighted by age, sex, education, and race using the latest available U.S. Census figures. The margin of error for each poll is noted above."

Leaving aside weighty questions about the assumptions the authors are making about their data that would make such weighting appropriate and whether or not the "margin of error" (whatever that might be) was correctly calculated, let us consider the phrase "weighted ...using the latest available U.S. Census figures." What a world of trust is embodied in that phrase! What an enormous amount of data in the private sector is buttressed by appeals to that phrase, for ABC News and the Washington Post are not alone in using Census data to adjust their sample data. All survey organizations who have any concern for bias in their samples do it. Indeed, the basic Census data are the foundation on which sample surveys, including the most widely used ones conducted by the Bureau itself, such as the Current Population Survey, are based. For example, NORC's Sampling Department uses various Census

Bureau products. These range from the special series printed reports from the CPS and Annual Housing Surveys to the 1980 Census tapes, maps, and printed reports. Data available in special reports and on CPS tapes are frequently used as reference material in survey and study design to obtain current population estimates. Census material, such as those from the 1980 Census, are used to develop national, local area, and special population frames.

Within the last year we have planned and implemented various studies that rely on the 1980 Census block data. Some designs have required only general variables, e.g., total population and total housing counts. Others have required more detailed counts for race, ethnicity, and country of origin at the block level and the PL-94 racial population counts available at the block level. We are currently redesigning our national sample frame based on the 1980 Census data. We have been delayed at least a year in accomplishing this redesign because of the delay in the release of the necessary Census information.

When the "latest available U.S. Census figures" are no longer very up-to-date, the adjustments that are made on the basis of those figures will not be doing thejob of providing the corrections that people who use them depend upon. As we move further from the year of the Census itself, we depend more and more heavily on data from the Current Population Survey for estimates of population and its distribution according to various characteristics. If the CPS it elf is not redesigned to take into account the 1980 Census or if its sample size is reduced so that the sampling error for various estimates is increased, the "latest available U.S. Census figures" becomes a very uncertain standard against which to calibrate private sample surveys.

The business of adjusting sample data to match population figures derived from Census data is an arcane matter of interest to few people and understoodby only a few specialists. Furthermore, it is a process that goes on

behind the scenes with little fanfare. Indeed, it is only recently that standards for reporting polls have been formulated so that as much information as I found in the ABC News release is routinely given in news releases—inadequate as it is for any serious evaluation of the method of conducting the poll. But lest we think these are trivial matters, consider the magnitude of the phenomenon. A recent report by the Panel on Survey Measurement of Subjective Phenomena of the Committee on National Statistics estimated that more than 20 million survey interviews are conducted in the United States each year. If there were an average sample size of 1,000 per survey, something like 20,000 surveys a year are being conducted. Decisions about marketing various products, the assignment of personnel to different tasks, the evaluation of programs, estimates of future expenditures, and a host of other decisions affecting the fate of government programs, corporate affairs, and individual lives are made on the basis of or deeply influenced by data collected from sample surveys. At some point in its collection or analysis most of these surveys use federal statistics, mostly from the Census Bureau, as a standard against which to measure their data and to adjust that data to bring them in line with the federal data. If the standard can no longer be trusted, what are we to make of all those other data?

It is not only that the functioning of government is hampered by the reduction in support for federal statistics, but, more importantly, the private sector itself—businesses, labor unions, private as well as public social agencies, public interest groups of all sorts. The bulk of the surveys are conducted by, paid for, and used by private concerns. Their decisions, as well as those of the federal government, are affected by the quality of the data.

Those involved in the political process have always been ambivalent about data. Statistics are used in a variety of ways to buttress political arguments either in support of one's own position or to attack that of one's

opponents. Ambiguous data make that process easier. But what if one is seriously trying to conduct policies so as to achieve a certain inflation rate or a certain unemployment rate or to manage the growth of money supply within fairly narrow limits? If you cannot trust the measures of those quantities, how can you know whether your policy is working or not?

Business executives are more concerned with results and less concerned with political positions as such. The much-loved "bottom line," whether it be conceived in short-run or long-run terms, governs their behavior in a competitive, profit-oriented economy. Accurate information is vital to the running of a business, to the scheduling of production, to making personnel decisions, to making decisions about expanding or contracting facilities, investing in new equipment or entering new markets. Much of the needed data for business decisions is about the environment in which the firms operate; much of it is about the economy and society in which they do business. The accuracy of these data is, in many subtle and often indirect ways, dependent on the accuracy of the federal statistics that form the standard against which much of their own data are compared.

I have talked so far almost exclusively in terms of the importance of federal statistics for sample surveys. Increasingly, however, data from surveys and other sources are used in complex models of the economy, economic forecasting models, energy systems models, etc. Such models have become important tools for policy analysis, allowing policymakers to look at the consequences of adopting different policies or at the effects of different combinations of policy outcomes (assuming, of course, that the model is a correct approximation of reality). For example, in her testimony before the Garcia Committee, Dr. Constance Citro, Vice President of Mathematica Policy Research Inc. wrote:

> "Mathematica Policy Research has developed a computer model of the U.S. population, the MATH model, that has been used extensively by Congressional and Executive branch agencies for cost-benefit analyses of current government tax and transfer programs and proposed alternatives.
>
> "For many studies, MPR analysts rely heavily on publicly-available data sets obtained from federal statistical agencies such as the U.S. Bureau of the Census, Bureau of Labor Statistics, Energy Information Administration and others. The computer model of household behavior...requires a large household data file as input. The model regularly uses the March CPS data files...and has also used public use microdata samples of the Decennial Census and the 1976 Survey of Income and Education. Needed data items that are not on these general household data sets are frequently obtained using statistical merging techniques with specialized data sets such as the Health Insurance Survey, the Annual Housing Survey, and the Survey of AFDC Participants." (Hearings, Series 97-41, p. 272)

Now complex models of the MATH variety have their own problems and depend for much of their validity on the assumptions of the mathematical relationships that go to make them up, but it is clear that no amount of mathematical sophistication or theoretical understanding of the interrelationships of variables can make up for fundamental errors in the input data. The maxim "garbage in, garbage out" needs to be pondered seriously. While we are not in imminent danger of seeing our major statistical series degenerate into garbage, we must use every effort to see that incipient spoilage does not develop and that the new products are at least as good as the old.

I have stressed the role of federal statistics as a standard against which other statistics are measured because I think this is an aspect of the debate about the effects of cuts in the statistical budget that is most often overlooked. The debate on the budget cuts has been almost entirely in terms of the direct effects of the cuts, particularly on statistical programs—reports that are no longer prepared, data collection programs that have been cancelled, etc., and how the curtailment of particular programs affects users

directly. The indirect effects are harder to trace and to make plain to nontechnicians--even, in many instances to sophisticated users because the assumption of the quality of federal statistics is so much a part of our experience that it is difficult to see all of the ways in which an erosion of quality would affect what we do.

Of course, specialists who work with (and work over) federal statistics all the time have long been worried about many aspects of the quality of some federal statistics. We have pushed, with some success until recently, to spend more money on improving statistics, not less. I say this not to criticize those responsible for the federal statistical system, for they are the ones most acutely aware of the shortcomings, but rather to point out that the problem is larger than just preventing an erosion of quality, important as that goal is at the present time. We must do more, not less, to strengthen and improve the system over what it was before the cuts--an even more difficult and challenging task!

While attention has been focused on the restoration of cuts for specific programs--attention that has been quite successful so far, largely due to the efforts of those with whom I share the platform today--we must not neglect other things that are going on which will have an impact over the long run, perhaps an even greater negative impact on the quality and usefulness of federal statistics as a standard than any specific programmatic cuts that have been contemplated. The most important of these is the abolition within OMB of the Statistical Policy Board and the position of the Chief Statistician. The erosion of quality control over federal statistics represented by the Statistical Policy Board (under its various names) has been going on for some time. It appears to be extremely difficult for people not well versed in statistical matters to take seriously the effects of the neglect and final

abandonment of that office. Unfortunately, one cannot show dramatically how someone died from a poisoned statistic or that thousands became unemployed because of faulty data. While folklore tells us that a kingdom was lost for want of a nail, we do not have a graphic image to tell us what can be lost for want of an accurate statistic. It is quite likely that policies have been adopted that have wasted millions of dollars or disrupted thousands of people's lives because the data available to the policymaker were faulty, but we cannot document it in the striking fashion that a good picture or a good anecdote can.

What are the ways in which the abolition of a central office of statistical policy may affect the quality and usefulness of federal statistics? One of the easiest to see if you have had any direct experience using statistics is an inevitable increase in the variance in statistical definitions. Even within one organization it is difficult to keep definitions or the questions asked to measure variables identical. Anyone who has wrestled with the problems of industry or occupational classification codes used in different data sets knows that small changes in the definition of classification schemes can wreak havoc on the comparability of data. One of the more striking examples of the effects of changes in classification criteria is seen in the 1980 Census population figures for people of Hispanic origin. How much of the reported increase in the size of the Hispanic population is due to real increase, how much is due to improved methods in conducting the Census that reduced the undercount and how much is due to changes in the way Hispanics were classified? We may never be able to disentangle the effects, certainly not until more experience is had with the new way of asking for ethnic identification. Coordination across agencies or even across separate divisions of the same agency is difficult enough when

there is a strong central office. Without it, no one who has the slightest experience with the real world of statistical programs can believe that there is any hope for anything but a gradual reduction in standardization and a growing diversity in definitions. We will increasingly have conflicting numbers due in large part to differing systems of classification.

Part of the reason that it is difficult for people to see the problem is that the process of generating the numbers in statistical reports is a mysterious one for many people. They fail to realize the extent to which counting things, particularly attributes such as ethnicity, income, occupation, employment status, etc. depends upon classification rules that are not all that obvious. While what constitutes a person may be clear, what constitutes a household is not; and what constitutes being unemployed depends on a number of factors, some of which are socially defined and may change over time. When classificatory categories are grouped, the rules for grouping are also subject to debate, and the rules may be more or less appropriate for different purposes. Indeed, for many classification and counting problems, the purpose of the measurement is very important in determining what is the appropriate classification scheme. Even something as seemingly simple as educational level can be defined in many different ways. One of the important functions of a central statistical authority is to keep the system of official statistics as standard as possible so that meaningful comparisons can be made between data sets and over time.

There are, of course, many other ways in which a central office affects the quality and usefulness of statistics: in exercising their review function over questionnaires, in examining the new paperwork budgets, in providing technical expertise and guidelines to agencies with less than adequate staffs, to name only a few. I have discussed the problem of

standardization of classification rules because I think it illustrates well not only the complexity of the standards problem but also the fact that the problems are often technical in nature and require a lot of time to solve. The fact that they require close attention and that dealing with them is time consuming implies that a sizeable staff is needed in the central office if the central function is to be performed adequately. I believe that the rather sanguine view of the OMB officials responsible for what remains of the central statistical function that they will be able to fulfill their function responsibly with a small staff comes from a failure to understand the nature of the tasks that are confronting them. The problem is not one of general policy and oversight that can be accomplished by a few talented individuals. The tasks require time and the close attention of a large staff if they are to keep up with the day-to-day requirements mandated by law and still have time to develop statistical policies which will lead to the improvement in the quality of federal statistics.

I would like to turn now to one final issue in the current debate, the role of the private sector in the federal statistical system. There are two issues here. One concerns who collects the data; the other, who pays for the data collection. In previous administrations, particularly Republican ones, the question of contracting out the data collection for many statistical programs had been raised. In fact, we have seen over the past decade the rise of a modest new industry of social researchers who provide a variety of statistical services, including data collection, for the federal government. In large measure we have in fact seen a shift toward the private sector, at least as far as actually performing much of the statistical work required by federal agencies. In the current debate, however, a new argument has emerged, namely, that since much of the data collected by the federal

government is of vital interest to the private sector, the private sector should provide the data for itself. While this argument has a superficial plausibility to it, at least for some data, it is based on a fundamental misunderstanding of the nature of the federal statistical system, and, ironically, on a misunderstanding of the nature of the free enterprise system.

For a product to be provided by the private sector in our system, there must be a market for the product, and the firm or firms providing the product must be able to control its production and sale in order to profit from supplying the product. The production of statistics of the sort, for example, that the Census Bureau produces requires the expenditure of very large sums which would be prohibitive for private firms even if they were able to do it at reduced cost. The price for the data that the producer would have to charge would depend on how many purchasers there were, but the producer would have to be confident that there would be enough purchasers willing to pay a price large enough to recover his costs and make a profit.

The problem, as I see it, is that the demand for the statistics produced by the system as a whole is quite large, but the market for the statistics of any particular survey is limited--so limited in fact that it is unlikely that private entrepreneurs would risk the capital necessary to develop and market any particular survey. Furthermore, even if producers judged that the market was there, the exploitation of the market would depend on their ability to control the data in such a way that buyers could not resell it to others at a cheaper price. This rediffusion problem has been a barrier to the development of data archives on a self-financing basis. More fundamentally, I believe that the nature of the market is such that it is unlikely that any firm would be induced to risk the costs of producing the

data. In addition, the evidence so far regarding the private market for data does not give me much confidence that the data demanded by that market is of very high quality. While it is possible that without the high quality data available at modest price from the federal statistical system, a new set of producers of high quality data would arise, it would take some time for that to happen and we would have to endure a number of years with poor or at best spotty data. Furthermore, the buyer who has the biggest need for high quality data is the U.S. government itself, who then might end up paying for it by buying it from the private supplier.

Statistics on society and the economy are better thought of as public goods necessary for the furtherance of the public interest rather than private goods serving private interests. Although private interests may also benefit from such public goods, it does not follow that they will, or should, pay the full costs of their production.

Let me try to summarize this rather rambling presentation. I have made four points:

1. Federal statistics form the foundation and are the standard for private statistics. Serious damage to the federal statistical system will have even greater negative effects on private statistics.

2. Decline in the quality of federal statistics will not only affect the accuracy of the large amounts of data collected through sample surveys by the private sector, but will also damage many of the forecasting and analytic tools based on various models of the economy or sectors of the economy, which are increasingly important tools for management of large business concerns, as well as policy planning in governmental bodies, because the efficacy of these models is limited by the quality of the input data.

3. It is the standards function that will suffer most with the cuts in funding and the virtual elimination of the central statistical group in OMB. Specific programs or surveys can be restored with some loss, but the loss of coordination in such matters as common definitions and common classificatory schemes will result in long-term impairment of the quality and usefulness of federal statistics. Enormous amounts of money would then have to be spent to repair the damage.

4. Finally, I have argued that statistics on society and the economy of the sort provided by the federal statistical system are public goods and are properly the responsibliity of the federal government rather than the private sector.

I close with a remark by the director of marketing research of one of the country's largest corporations during a meeting with some congressional staff members about the damage the budget cuts might do to the Census Bureau's programs. When asked about the importance of Census Bureau data to his firm, he replied, "We regard data from the Census Bureau as the God-given truth." While I would not go so far, I think the remark does reflect the role that federal data more generally plays as the foundation upon which data in the private sector are built. If cracks develop in that foundation, we had better develop some good props to hold things together or run for cover.